Paper prepared for the 122nd EAAE Seminar
"EVIDENCE-BASED AGRICULTURAL AND RURAL POLICY MAKING: METHODOLOGICAL AND EMPIRICAL CHALLENGES OF POLICY EVALUATION"
Ancona, February 17-18, 2011

Methodological innovations in estimating the (inverse) relationship between farm productivity and farm size in a developing economy: a case study of Burundi

Verschelde M.[1], Vandamme E.[2], D'Haese M.[2] and Rayp G.[1]

1 Ghent University/Department of general economics, SHERPPA, Ghent, Belgium
2 Ghent University/Department of agricultural economics, Ghent, Belgium

Marijn.Verschelde@ugent.be

Methodological innovations in estimating the (inverse) relationship between farm productivity and farm size in a developing economy: a case study of Burundi

Marijn Verschelde Ellen Vandamme Marijke D'Haese
Glenn Rayp

Ancona - 122$^{\text{nd}}$ EAAE Seminar
"Evidence-Based Agricultural and Rural Policy Making"

Abstract

We use a nonparametric estimation of the production function to investigate the relationship between farm productivity and farming scale in poor smallholder agricultural systems in the north of Burundi. Burundi is one of the poorest countries in the world, with a predominant small scale subsistence farming sector. A Kernel regression is used on data of mixed cropping systems to study the determinants of production including different factors that have been identified in literature as missing variables in the testing of the inverse relationship such as soil quality, location and household heterogeneity. Household data on farm activities and crop production was gathered among 640 households in 2007 in two Northern provinces of Burundi. Four production models were specified each with different control variables. For the relatively small farms, we find clear evidence of an inverse relationship. The relatively large farms show a different pattern. Returns to scale are found to be farm scale dependent. Parametric Cobb-Douglass models tend to over-simplify the debate on returns to scale because of not accounting for the different effects of large farms. Other factors that significantly positively affect production include the soil quality and production orientation towards banana or cash crop production. Production seems to be negatively affected by field fragmentation.

Keywords: inverse relationship, farm size, nonparametric, Burundi

JEL classification: D24, O13, Q12, Q18

1 Introduction

Burundi has the sad record of being one of the poorest countries in the world. With a GDP of 380$ (PPP) per capita it is ranked at the bottom of the group of low-income countries (WorldBank, 2009). In the Human Development Index ranking of 182 countries, it is at the 174th place. The country seems to have much against it when trying to succeed in promoting economic growth; its size is rather small, it is landlocked, with limited natural resources and it is prone to ethnic conflict. The economy depends largely on agriculture; more than one third of the total GDP is derived from agricultural production and more than 90% of employment is allocated to the agricultural sector. Agriculture also plays a vital role in the trade balance as more than 90% of foreign exchange earnings is derived from the export of coffee although the contribution of this export to the country's GDP is rather small (CIA, 2010).

Burundian population has been booming[1] with far-reaching consequences for natural resources (Cochet, 2004) and political stability (Bundervoet, 2009). On the one hand, population growth in sub-Saharan Africa increases the pressure on agriculture as more mouths need to be fed. Currently FAO categorizes Burundi as a low-income, food deficit, country. More than half of the population (63%) suffers from undernourishment (FAOSTAT, 2005) and many more are food insecure. This is a clear indication and warning that the agricultural sector is not up to the challenge of feeding the local population. On the other hand population growth contributes to decreasing average farm size due to subdivision at heritage and fragmentation of smaller plots within farms (Jayne et al., 2003). The decrease in access to land and the small plot sizes have implications for the farming system in general and farming strategies in particular. This leads us to the question of impact of farm size on production levels. In case the farming sector is vulnerable and unable to meet the needs of the growing population, it may, according to Malthus (1798) lead to negative checks[2].

[1] According to the CIA World Factbook (consulted August 2010) the estimated population growth rate in Burundi is 3.67%. The mean age of the population is 16.8 years, with 46.3% of the people being less than 14 years old. Life expectancy is 57.8 years. Only 10% of the population lives in urban areas.

[2] Malthus' point was that as humans "reproduce" they continually put pressure on the resources for subsistence, which eventually is halted by checks to population growth such as war and epidemics (Leathers and Foster, 2009): "land, unlike people, does not breed" (paraphrasing Malthus in Leathers and Foster (2009)); or in other words, if malnutrition or ill-health become too problematic due to a lack of subsistence means, the risk on a population correction such as war increases (Leathers and Foster, 2009).

Conversely, Boserup (1965) argued that pressure on the agricultural system induces innovation, leading to an increase in agricultural productivity and as such, in food production (Boserup, 1965).

Literature points to the utmost importance of increasing land and labour productivity in the agricultural sector in order to achieve an increase of the African food production (Collier and Dercon, 2009; Piesse and Thirtle, 2010). The (possibly inverse) relationship between farm/plot size and land productivity has been heavily debated over decades now (see references section 2; see also the introduction by Wiggins et al. (2010) of a special section in the November 2010 issue of World Development on the future of small farming). In particular, Collier and Dercon (2009) point to the need for increasing labour productivity on African smallholder farms. Agricultural labour productivity in small-scale farm systems is found to be very low, this is mainly due to the reported overallocation of (family) labour also referred to as hyper-exploitation of family labour, which is basically a problem of very low marginal labour productivity levels (Barrett, 1996).

Important policy issues that emerge are not only how productivity could be increased, but also whether the focus on small — family oriented — farms is the right vehicle for achieving productivity growth. Since Schultz (1964) small farms are considered to be efficient in what they do (Schultz, 1964), and support has been geared towards these smallholder producers. Yet, are they up to the challenge of feeding the growing population? (Wiggins et al., 2010) Are they currently productive enough to meet increasing food demand in the future? The contribution of our study to these questions is that we analyze the factors influencing productivity using a non-parametric estimation of the production function estimated for a unique dataset in the North of Burundi. The results point to diminishing but positive returns to scale. The relationship between inputs and farm output – here measured as market value of crop and coffee production – is not linear, which parametric models fail to capture. The next section explains which gap in literature we want to help filling.

2 Research background

According to neoclassical economics, optimal production levels are reached when marginal productivities are maximized. Perfect factor markets ensure an optimal allocation of different production factors which will lead to these maximal marginal productivities. When applying this theory to farming it implies that inputs and production factors such as land, labour and capital are allocated in such a way that yields (output per land unit) and productivity (output/input) are maximal and virtually equal for all farms. Yet, factor markets are failing in developing countries, and the transaction costs farmers need to incur in order to reach input and output markets are significant. This partly explains why an IR between size of production and productivity that is found in several developing areas (Lipton, 2010) contradicting the theories of economies of scale.

Several obvious and less obvious reasons and explanations for this IR have been tested and proven. A primary obvious reason is the presence of imperfect factor markets (Feder, 1985). This includes failures in different types of production factor markets: land market (Platteau, 1996; Heltberg, 1998), credit market (Assunção and Ghatak, 2003), insurance market (Dercon and Krishnan, 1996) and labour market (Feder, 1985; Barrett, 1996; Assunção and Braido, 2007). Malfunctioning or a complete absence of these markets will lead to suboptimal resource allocation on farm level implying inefficiencies. An important cause of the presence of imperfect labour markets in developing countries is claimed to be labour supervision cost (Feder, 1985; Lipton, 2010). The theory of imperfect labour supervision claims that labour productivity of family labour forces is higher than of hired external labour forces. As hired labour is less motivated and effective, it takes more productive family labour to supervise hired labour which decreases overall labour productivity at farm level. This would explain why labour and farm productivity are lower on large farms, which require more hired labour. A second important explanation is related to farm management. Farming practices and production methods might vary according to farm size, leading to differences in yields and productivity (Byiringiro and Reardon, 1996; Schultz, 1964; Assunção and Braido, 2007; Lipton, 2010).

A third explanation of the IR is related to methodological issues. Recent research questions whether the IR between farm size and productivity emerges (or not) due to omitted variables. Soil quality is mentioned as an important but often neglected explanatory variable. Differences in soil quality lead to differences in soil productivity which clearly affect output (Sen, 1975), with small farmers being more productive because of having plots of better quality. All revised studies on this issue show a decrease in the severity of the IR when controlling for soil quality (Lamb, 2003; Assunção and Braido, 2007; Barrett et al., 2010). Benjamin (1995) finds that the IR disappears when indirectly controlling for soil quality (Benjamin, 1995). A second set of missing variables are household specific characteristics such as household size, dependency ratio, and gender of the household head (Assunção and Braido, 2007; Barrett et al., 2010). However none of the studies cited up to now has proven household characteristics to solely explain the IR. Moreover, Lipton (2010) argues that differentiation in farm management skills (as mentioned above) as an explanatory variable of farm productivity was not yet sufficiently tested in empirical research.

Consequences of the presence of this IR are quite far reaching. If small farms tend to be more efficient in developing countries, supporting these small-scale farmers is the way forward. However, as mentioned above, literature reports on methodological problems in proving the IR (Lipton, 2010). First, it is important to acknowledge the presence of explanatory omitted variables. Secondly, most empirical studies on the IR are based on cross sectional data. Arguably, the scale ranges on which the analysis are based is too small to measure scale effects. Analyses will compare the smaller farmers with the less-small small farms, and fail to measure a longitudinal effect of scale increase (Collier and Dercon, 2009). Another methodological issue is on distinguishing between small-ness and family-ness (Lipton, 2010). As mentioned above, a popular explanation for the inverse productivity relationship is the cheaper labour supervision of family labour. Hence the IR could be seen as an indication that not the size of the farm matters, but the family control over its production decisions.

In this paper we try to address a number of important empirical issues. First, we account for mixed output by calculating the market values of all crops produced while allowing for mixed cropping systems. Secondly, by using a nonparametric approach we are able to track heterogeneity in productivity effects of increased access to production factors Thirdly, our rich dataset allows controlling for several of the missing variables mentioned above. The data collection and methodology is explained in the next section.

3 Methodology

3.1 Data

Household data on farm activities was gathered in 2007 in two densely populated provinces of in the North of Burundi, Ngozi and Muyinga. The provinces were chosen because they are among the most populated of the country. Both provinces cover an area of 2300 km^2 and 1.4 million inhabitants; this is 13% of the total surface of Burundi and 19% of the population. Both provinces are densely populated with 475 inhabitant per km^2 in Ngozi and 322 inhabitants km^2 in Muyinga. Economic activity outside agriculture is very limited in both provinces, except for the city of Ngozi which is the third largest city of Burundi. In total 640 farm households were questioned; 360 in the Nogzi Province and 280 in Muyinga Province. All 16 municipalities of the two provinces were covered (nine in Ngozi Province and seven in Muyinga), per province ten villages where selected based on geographical distribution and in every village four households were randomly selected. The interviews were held in Kirundi in collaboration with a team of the University of Burundi. Because of missing data, 20 farms had to be excluded from the data analysis.

For each household, two questionnaires were used; a first questionnaire collected information on household and farm characteristics. A second questionnaire was used to gather information on each plot the farmer owned. The result is a very rich dataset with detailed and reliable information on farm scale (production level, size, labour input, farm inputs), the farming system such as crop choices and cash crops as well as on the farmer's evaluation of the soil quality, and steepness of the different fields. The latter is particularly important given the area is particularly hilly.

3.2 Variables included in the model

The output is measured by the sum of the market value of all crops produced irrespective of whether these are sold or consumed by the household. Farm production for each food crop is multiplied by the average market price of the respective crops. The level of marketing by the farmers is so low that no individual farm-gate prices could be captured. Furthermore, the diversity of the mixed cropping produce made it not possible to use other quantities. The alternative of caloric content was also not used because it would not be possible to account for the value of coffee production.

Factors influencing production are production factors (land, labour, inputs), while controlling for location, farm management, soil quality and household characteristics. As land input, the farm area that is actually used for cultivating food and cash crops is included. Two different sources of labour are distinguished, namely family labour (expressed in person units) and hired labour (expressed in paid wages). One other type of non-labour inputs is included: the sum of the expenditure on seed, chemicals and agricultural equipment.

Four different types of control variables are included: location, farm management, soil quality and household heterogeneity. Location is considered by adding a dummy for the

province.
As the capital of the Ngozi province is the third largest city in Burundi, access to assets and markets in this province might be significantly higher than in Muyinga. Indicators for farm management are the cropping pattern, fragmentation index and production technology used. A mixed cropping pattern is quantified by the share of the total cropping surface used for either: staple crops, cash crops, banana or other crops. Land fragmentation is assessed by the Simpson index. This index varies from zero to one and is calculated by dividing the total sum of the different field surfaces squared by the square of total cropping area ($S = \sum s_i^2 / \left(\sum s_i\right)^2$). Farms with higher land fragmentation will demonstrate a higher Simpson index. Two dummies are included to account for the use of chemicals and animal manure as soil improving farming techniques. Farmers were asked to assess the steepness of the plot and soil quality of each of their plots on a scale from one to four. This resulted in the calculation of two variables, one variable that indicates the share of the total cropping surface that has a steep slope and a second variable representing the share of the total cropping surface with good quality soil.
Finally, we control for household heterogeneity by including the following variables: age of the household head, the share of household income derived from off-farm activities and a dummy for extension (whether or not the household has been visited by an extension officer). A descriptive analysis for all variables included in the model is given in Table 1.

Variables	Ngozi province		Muyinga province		Entire sample		Test
							t-test
Agricultural output (1,000BIF)	1029.67	(1062.04)	787.60	(948.41)	921.13	(1,019.01)	2.99**
Farm size (ha)	9.87	(1.44)	1.29	(1.89)	1.13	(1.66)	-2.26**
Farm size per person (ha/pers)	0.18	(0.24)	0.25	(0.35)	0.21	(0.29)	-2.68**
Size cultivated land (ha)	0.65	(1.1)	0.87	(1.1)	0.75	(1.11)	-2.44**
Size cultivated land per person (ha/pers)	0.12	(0.19)	0.17	(0.18)	0.14	(0.19)	-3.25**
Family labour (nb)	2.74	(1.34)	2.51	(1.10)	2.64	(1.24)	2.30**
Labour cost (paid wage, 1,000BIF)	39.34	(13.66)	23.91	(100.77)	32.42	(118.35)	1.66**
Cost for seeds (1,000BIF)	20.46	(34.00)	17.62	(20.70)	19.18	(28.82)	1.28
Costs for chemicals (1,000BIF)	8.45	(20.56)	1.10	(5.98)	5.16	(16.19)	6.29**
Costs for agricultural material (1,000BIF)	4.47	(9.65)	3.76	(6.87)	4.15	(8.52)	1.02
Total cost production inputs (1,000BIF)	33.38	(48.38)	22.49	(25.00)	28.49	(39.98)	3.61**
Share staple crops (%)	52.51	(19.57)	61.88	(18.81)	56.71	(19.78)	-6.04**
Share coffee (%)	13.77	(13.62)	9.22	(10.71)	11.73	(12.60)	4.65**
Share banana (%)	20.78	(14.60)	18.05	(12.29)	19.55	(13.67)	2.53**
Share non-productive land use (%)	12.93	(17.27)	10.84	(17.02)	11.99	(17.18)	1.52
Share in the marsh (%)	9.33	(12.28)	2.87	(6.29)	6.40	(10.54)	8.46**
Share under steep slope (%)	20.52	(29.85)	17.57	(29.59)	19.20	(29.75)	1.23
Share good quality soil (%)	49.51	(37.53)	46.49	(41.43)	48.15	(39.32)	0.94
Fragmentation index (0-1)	0.23	(0.14)	0.24	(0.14)	0.24	(0.14)	-0.51
Age of hhhead (years)	41.36	(12.41)	40.01	(12.89)	40.75	(12.64)	1.32
Share income off-farm (%)	37.45	(3.59)	39.16	(32.04)	38.22	(32.33)	-0.65
							χ^2-test
Use of chemicals (% yes)	83		65		75		26.27**
Use of animal manure (% yes)	61		49		56		9.78**
Extension visit (% yes)	21		57		37		82.62**
Observations	342		278		620		

Significance levels : $*$: 5% $**$: 1% $***$: 0.1%

Table 1: Descriptive analysis dependent, independent and control variables included in model

3.3 Nonparametric regression approach

The empirical model is defined by a $n \times 1$ dependent scalar y, a multivariate regressor x and additive error ϵ.

$$y = g(x) + \epsilon \tag{1}$$

This production function can be estimated by imposing a parametric form. The vast majority of papers impose a Cobb-Douglass (CD) specification. Log output is defined as a linear function of the log of the q regressors, with additive error.

$$\ln y = \alpha + \sum_{k=1}^{q} \beta_k \ln x_k + \epsilon \tag{2}$$

However, if there are non-linearities or interactions in the true model, the empirical model is misspecified and coefficients are inconsistent (Henderson and Kumbhakar, 2006). A flexible parametric alternative is the Translog specification; quadratic effects and interaction effects are introduced in the empirical model.

$$\ln\, y = \alpha + \sum_{k=1}^{q} \beta_k \ln\, x_k + 0.5 \sum_{k=1}^{q} \sum_{l=1}^{q} \beta_{kl} \ln\, x_k \ln\, x_l + \epsilon \tag{3}$$

In some cases, the Translog specification can give economically unreasonable estimates, caused by (1) failure to capture all nonlinearities in the true model (Henderson and Kumbhakar, 2006), (2) the high multicollinearity or low degrees of freedom as result of the inclusion of quadratic effects and interactions.

To avoid imposing *'a priori'* a functional relationship between the output scalar and regressors, nonparametric approaches can be used[3]. In a nonparametric (generalized) kernel regression, $E[Y|X = x]$ is estimated by locally averaging those values of the dependent variable which have similar levels of the regressors (one could note it as $\hat{g}(x) = E[Y|\text{X close to x}]$).

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i w_i \tag{4}$$

We use Racine and Li (2004) generalized kernel weights to specify the weight function w_i for $x = [x^c, x^o, x^u]$, where x^c is a vector of continuous values, x^u is a vector of unordered discrete values, x^o is a vector of ordered discrete values. Kernel functions (l^c, l^o, l^u) are used to be able to give more weight to observations near the observation point. Window widths $(\lambda^c, \lambda^o, \lambda^u)$ impose the window of local averaging. If the window width is large, the curve will be a smooth straight line (as in a linear regression). On the other hand, if the window width is small, non-linearities are allowed for and the curve becomes less smooth.[4] It is intuitively clear and shown in literature that the choice of weighting function is of far less importance than the choice of the window of localization - which we will discuss below.

To construct the weight function for the local averaging, we specify a standard normal kernel function l^c to weight the continuous variables x^c. An Aitchision and Aitken (1976)

[3]See Li and Racine (2007) for an extensive overview of the used kernel regression approach

[4]Note the trade-off between bias and variance

kernel l^u is specified to weight discrete unordered variables x^u (see (5)). To weight the ordered discrete values x^o, we use a Wang and van Ryzin (1981) kernel function (see (6)).

$$l(X^u_{il}, x^u_l, \lambda^u_l) = \begin{cases} 1 \text{ if } X^u_{il} = x^u_l, \\ \lambda_s \text{ otherwise} \end{cases} \tag{5}$$

$$l(X^o_{im}, x^o_m, \lambda^o_m) = \begin{cases} 1 \text{ if } X^o_{im} = x^o_m, \\ (\lambda^o_m)^{|X^o_{im} - x^o_m|} \text{ otherwise} \end{cases} \tag{6}$$

To allow for a multivariate regression, we use - as is common practice - product kernels. The product kernel of X^c is $W_{\lambda^c}(X^c_i, x^c) = \prod_{k=1}^q (\lambda^c_k)^{-1} K((X^c_{ik} - X^c_k)/\lambda^c_k)$. For X^u, the product kernel is defined as $L_{\lambda^u}(X^u_i, x^u) = \prod_{l=1}^r l^u(X^u_{il}, x^u_l, \lambda^u_l)$. The product kernel of X^o is $L_{\lambda^o}(X^o_i, x^o) = \prod_{m=1}^s l^o(X^o_{im}, x^o_m, \lambda^o_m)$. All together, we can specify a Li-Racine generalized kernel function as $\mathcal{K}_\gamma(X^c_i, X^o_i, X^u_i) = W_{\lambda^c}(X^c_i, x^c) L_{\lambda^u}(X^u_i, x^u) L_{\lambda^o}(X^o_i, x^o)$, with $\gamma = (\lambda^c, \lambda^u, \lambda^o)$.

We estimate $E(Y|X = x)$ by the use of a local-linear estimator. The local-constant (Nadaraya-Watson) estimator takes the kernel weighted average of the observed y_i values and normalizes it by the sum of the kernel weighted averages (see (7)). This is the so called local-constant approach as it specifies a locally averaged constant value y for each observation point. It can be obtained as the solution of a in (8). The local-linear estimator estimates a local linear relation for each observation point by obtaining a and b in (9). If bandwidths are very large and there is thus no local weighting, we have the parametric least squares estimator. The least squares estimator can thus be seen as a special case of the local linear estimator (Li and Racine, 2007, p. 83). We opt for the local-linear regression as it has better boundary properties than the local-constant regression (Hall et al., 2007) and nests least squares as a special case.

$$\hat{g}(x) = \frac{\sum_{i=1}^n Y_i \mathcal{K}_\gamma(x, X_i)}{\sum_{i=1}^n \mathcal{K}_\gamma(x, X_i)} \tag{7}$$

$$\min_a \sum_{i=1}^n (Y_i - a)^2 \mathcal{K}_\gamma(x, X_i) \tag{8}$$

$$\min_{\{a,b\}} \sum_{i=1}^n (Y_i - a - (X_i - x)'b)^2 \mathcal{K}_\gamma(x, X_i) \tag{9}$$

As discussed, the choice of multivariate bandwidth γ is of crucial importance. We opt for the often used data-driven approach that minimizes the asymptotic integrated mean squared error (AIMSE): the least-squares cross-validation approach as defined in (10).

$$CV(\gamma) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{g}_{-i}(X_i))^2 t(X_i) \tag{10}$$

where $\hat{g}_{-i}$ is the leave-one-out local-linear kernel estimator of $E(Y_i|X_i)$, and $0 \leq t(\cdot) \leq 1$ is a weight function that serves to avoid difficulties caused by dividing by 0 or by the slower convergence rate arising when X_i lies near the boundary of the support of X. Simulation results of Li and Racine (2004) show that cross-validated local linear regressions indeed choose much larger bandwidths if the true relationship is linear.[5]

4 Results

4.1 Description of the farming system related to farm size

The farming system in Burundi consists of small peasant landholdings (of generally less than 1 ha per family as illustrated in Figure 1), very small plots with double cropping, manual self-subsistence farming with little marketed surplus (Cochet, 2004). Crop production is done on both the hill side and in the drained marshes. Two distinct cropping systems were distinguished on each landholding. A first system consisted of separate plots cultivated with mixed crops (grains, pulses, tubers and coffee), and, a second system was based on banana production (Cochet, 2004). Several authors emphasize the importance of banana production in the current farming system (Rishirumuhirwa and Roose, 1998; Cochet, 2004). It seems as if the banana has over the years replaced cattle production which requires more land and other natural resources. The most important food crops produced and consumed in the study area were sweet potatoes, beans, cassava, banana and flour of maize (FAO STAT, country profile, 2005). Except for banana and coffee, most farmers did not market produce and even when they did sell, it was mainly surplus sales of very small quantities.

The average farm size in our sample was 1.12ha however about 45% of the farms in the sample were smaller than 0.5ha. Farms were larger in Muyinga compared to the more

[5]We opt for this approach over the AIC CV approach as the least-squares CV approach is more used in the literature and is faster to compute.

densely populated Ngozi Province (see Table 1). The distribution of land over the sample was rather unequal. Moreover, compared to a previous study we find an increase of inequality in access to land, which resulted in an increased number of very small scale farms (smaller than 0.5ha) (Rishirumuhirwa and Roose, 1998). Furthermore farms were highly fragmented with on average more than eight plots on the hillside (collines), and one to two plots in the swamps (marsch). It is worth noting that the relatively large farms in our sample are deliberately not excluded from the analysis as they may contain valuable information which can be studied separately with a nonparametric model.

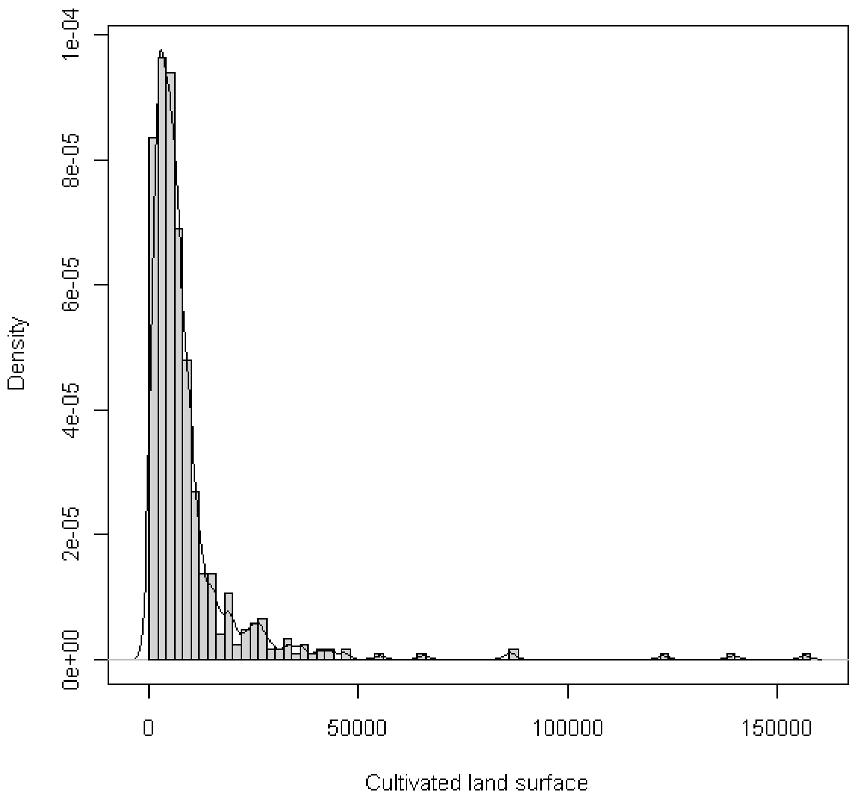


Figure 1: Density plot of farm sizes in the sample (m^2)

Symptomatic for the very poor livelihoods of the farm households in the study area, was their high level of food insecurity; the survey registered the HFIAS score (Household Food Insecurity Access Score, USAID, Coates et al., 2007), and results showed that 7% of the households could be considered food secure (results not shown in the table). Two thirds of all households interviewed were even labelled severely food insecure. These figures coincide with FAO data indicating that 68% of the total population is undernourished (FAO, 2009). Results presented in Table 2 suggest that farm size, production strategy, crop productivity and farm production may be related (to be analyzed next), although not all effects tend to go into the same direction. Large farms showed slightly different land use patterns compared to small farms. Larger farms tended to attribute a larger share of their total farm surface to other non-production activities such as forestry and fallow land whereas small farms used most of their land for staple food production rather intensively. However, the share of production area dedicated to cash crops, i.e. coffee production, did not significantly differ according to farm size quartiles. Small farms were using a larger proportion of the total production surface for banana production while larger farms used relatively more land for bean production (not detailed in the table). Farm proportions dedicated to other important crops in the area such as tubers and cereals did not differ be tween the land size quartiles and are therefore not reported. Crop diversification seems to be larger on larger farms making them less prone to risks of crop failure compared to small less diversified farms.

Variables	First quartile		Second quartile		Third quartile		Fourth quartile		Test
									F-stat
Agricultural output (1,000BIF)	398.46	(316.72)	557.57	(373.55)	804.62	(531.08)	1621.13	(1444.06)	72.02**
Farm size (ha)	0.20	(0.085)	0.51	(0.1)	0.91	(0.15)	2.93	(2.59)	139.94**
Farm size per person (ha/pers)	0.06	(0.04)	0.11	(0.07)	0.18	(0.11)	0.49	(0.48)	96.808**
Size cultivated land (ha)	0.15	(0.07)	0.38	(0.13)	0.66	(0.19)	1.8	(1.8)	101.69**
Size cultivated land per person(ha/pers)	0.04	(0.03)	0.08	(0.06)	0.13	(0.09)	0.30	(0.30)	79.98**
Family labour (nb)	2.18	(0.69)	2.54	(1.16)	2.82	(1.40)	3.01	(1.44)	13.77**
Labour cost (paid wage, 1,000BIF)	5.94	(27.27)	9.66	(20.96)	20.96	(43.02)	94.83	(221.39)	20.97**
Seed cost (1,000BIF)	11.95	(20.44)	15.71	(17.41)	19.73	(20.44)	29.63	(45.40)	11.22**
Costs for chemicals (1,000BIF)	1.06	(2.59)	3.49	(10.87)	4.22	(13.39)	12.03	(26.34)	14.00**
Costs for material (1,000BIF)	2.23	(3.11)	3.11	(4.02)	4.36	(7.10)	6.99	(14.37)	9.45**
Total cost inputs (1,000BIF)	15.25	(21.62)	22.31	(25.35)	26.26	(25.95)	45.44	(58.82)	21.83**
Labour cost per ha (1,000BIF/ha)	24.91	(105.6)	19.86	(47.21)	23.35	(47.81)	36.43	(101.53)	1.29
Seed cost per ha (1,000BIF/ha)	70.39	(115.96)	31.10	(34.20)	22.52	(23.55)	14.79	(29.33)	23.50**
Costs chemicals per ha (1,000BIF/ha)	6.13	(17.43)	6.85	(21.21)	4.74	(15.77)	5.55	(13.16)	0.42
Costs material per ha (1,000BIF/ha)	14.16	(23.55)	6.45	(9.01)	5.06	(8.55)	3.23	(7.04)	18.79**
Total cost inputs per ha (1,000BIF/ha)	90.70	(131.17)	44.41	(49.99)	32.33	(33.03)	23.57	(39.17)	24.64**
Share staple crops (%)	56.81	(20.31)	58.49	(19.55)	59.74	(18.34)	51.69	(20.01)	4.98**
Share coffee (%)	12.63	(14.68)	12.72	(12.13)	10.50	(11.40)	11.05	(11.90)	1.23
Share of banana (%)	23.97	(15.30)	19.72	(13.91)	18.49	(11.67)	15.98	(12.35)	9.55**
Share of non-productive land use (%)	6.59	(13.11)	9.08	(14.63)	11.27	(15.14)	21.28	(21.25)	23.93**
Share in the marsh (%)	8.63	(14.12)	5.92	(8.63)	5.43	(9.11)	5.76	(9.15)	3.06**
Share under steep slope (%)	19.46	(30.65)	16.41	(27.54)	19.30	(29.86)	22.77	(30.82)	1.23
Share good quality soil (%)	44.24	(39.28)	40.65	(38.22)	50.85	(39.73)	57.27	(38.35)	5.53**
Fragmentation index (%)	0.30	(0.17)	0.23	(0.12)	0.21	(0.10)	0.19	(0.14)	19.48**
Age of hhhead (years)	36.34	(11.15)	41.23	(13.21)	41.03	(12.61)	44.48	(12.24)	11.34**
Share income off-farm (%)	44.25	(33.16)	41.66	(34.10)	37.07	(30.66)	29.70	(29.52)	6.11**
									χ^2-test
Use of chemicals (% yes)	63.2		74.2		76.0		85.5		20.37**
Use of manure (% yes)	40.6		52.8		59.7		69.1		26.77**
Extension visit (% yes)	25.8		35.8		43.5		44.1		14.43**
Observations	155		159		154		152		

Significance levels : $*$: 5% $**$: 1% $***$: 0.1%

Table 2: Descriptive analysis for different quartiles of farm size (N=620)

The allocation of labour seems to be closely related to farm size with larger farms allocating more family labour and spending more money on extra labour. However, the level of labour per land unit was significantly higher for smaller farms as family labour per land unit was larger for small farms and wages paid for hired labour per land unit were not higher for larger farms. Investments in agricultural production were measured by the expenditure on seed, agricultural material and chemicals. These investments increased significantly with increasing production area. Smaller farms spent significantly more money per land unit on seed and agricultural material. Investments in chemicals such as fertilizer and pesticides were not different across the land size quartiles; these chemicals were used with the same, generally very low, intensity on both small and large farms. However the likelihood of using chemicals was larger on larger farms. On top of this, the likelihood of using specific soil improving techniques (manure, compost, mulching) was higher for the quartile with the largest farms. These findings suggest differences in the production strategies related to differences in cropping area. These differences in crop production strategies might lead to different production outcomes and even more so to differences in farm productivity.

4.2 Parametric approach

We start the estimation of the production model with the Cobb-Douglass approach. As there is too few variation in family labour, it is dangerous to consider this as a continuous variable. We define family labour as an ordered discrete variable. In contrast to nonparametric models, an ordered discrete variables cannot be included as one variable in a parametric model. Dummies are needed. To avoid multicollinearity, we limit the order from 10 to 3 by defining a dummy for family labour if the value is 3 or 4 and a dummy if the value is larger than 4. Farms that use no hired labour or intermediary inputs are not excluded. We include the dummies "Use of hired labour" and "Use of intermediary inputs" which equal 1 if used. As shown in Table 3, the four inputs (land use, family labour, hired labour and intermediary inputs) are found to have a positive and significant effect on output. However, the effect of increasing family labour from 1 or 2 to 3 or 4 persons was only significant at the 10% confidence level. We find no positive effect of increasing family labour above 4 persons. The fixed effect for province was significant with a higher output in the Ngozi province. In addition, Table 3 shows clearly that the output elasticity for cultivated farm area was smaller than 1. There is thus an IR found between farm size and farm output per unit of land. As the sum of output elasticities of the regressors is significantly lower than 1, the Cobb-Douglass model finds diminishing returns to scale. However, as noted in Section 2, the Cobb-Douglass does not allow for quadratic effects and interactions between the log of the regressors.

To introduce interactions and quadratic effects, we test the proper working of the Translog model for this data set. Results for the Translog model are summarized in Table 4. Surprisingly, we find no significant effect any more from the inputs the farmers used. We only find a significant quadratic effect of cost of labour and a significant interaction effect between cost of labour and cost of intermediates. As these results are in sharp contrast to the Cobb-Douglass model, we have doubts on the value of these results. The variation in the model is too low to include all the quadratic and interactions effects. Instead of an iterative process of step-wise reduction of the parametric Translog model, we opt for an alternative approach: the nonparametric regression as described in Section 2.

	Estimate	Std. Error	t-value	p-value
(Intercept)	9.53	0.32	30.20	0.00***
Log cultivated land	0.40	0.03	11.62	0.00***
Family labour: 3-4	0.13	0.08	1.72	0.09°
Family labour: 5 or more	-0.05	0.06	-0.87	0.39
Log hired labour cost	0.17	0.03	5.68	0.00***
Log costs intermediary inputs	0.07	0.03	2.31	0.02*
Province	-0.29	0.06	-5.08	0.00***
Use of hired labour	-1.33	0.29	-4.57	0.00***
Use of intermediary inputs	-0.33	0.31	-1.07	0.29
Adjusted R^2	0.47			
Observations	620			

Significance levels : ○: 10% ∗ : 5% ∗∗ : 1% ∗ ∗ ∗ : 0.1%

Table 3: Cobb Douglass Model

	Estimate	Std. Error	t-value	p-value
(Intercept)	9.99	1.84	5.43	0.00***
Log cultivated land	0.25	0.41	0.62	0.54
Family labour: 3-4	0.13	0.08	1.64	0.10
Family labour: 5 or more	-0.05	0.06	-0.86	0.39
Log hired labour cost	-0.19	0.31	-0.63	0.53
Log costs intermediary inputs	0.12	0.33	0.35	0.72
Province	-0.29	0.06	-4.99	0.00***
Log cultivated land^2	0.01	0.02	0.29	0.77
Log hired labour cost^2	0.03	0.02	1.68	0.09°
Log costs intermediary inputs ^2	-0.00	0.02	-0.03	0.98
Use of hired labour	0.88	1.49	0.59	0.55
Use of intermediary inputs	-0.61	1.49	-0.41	0.68
Log cultivated land × Log hired labour cost	-0.00	0.01	-0.39	0.70
Log cultivated land × Log costs intermed. inputs	0.00	0.01	0.17	0.87
Log hired labour cost × Log costs intermed. inputs	-0.01	0.00	-2.97	0.00***
Adjusted R^2	0.48			
Observations	620			

Significance levels : ◦: 10% $*$: 5% $**$: 1% $***$: 0.1%

Table 4: Translog model

4.3 Nonparametric approach

The nonparametric approach makes no *'a priori'* assumptions on the functional relationship between the dependent variable and regressors. Using cross-validation, the trade-off between bias (for a given model, larger for a smooth, linear curve) and variance (larger for a wiggly, non-linear curve) is settled. We illustrate the nonparametric results by showing directly the estimated level of output as a function of the value of a respective independent variable, holding the other regressors equal to respectively the median or modus. In addition, we show 95% bootstrap confidence intervals. A significantly increasing (resp. decreasing) curve illustrates a significant positive (resp. negative) effect of the regressor on agrictultural production.[6] As we did not find any effects for the dummies for use of hired labour and use of intermediary inputs, the dummies were excluded from the nonparametric model.

The base model includes as independent variables, size land used for agricultural production, family labour, cost of hired labour, cost of inputs used, and a dummy for the province (see Figure 2).[7] The model shows significant effects of cultivated land and cost of hired labour. The model confirms that production was higher in Ngozi compared to Muyinga. An increase in family labour did not significantly contribute to production, indicating a very low (zero) marginal productivity of family labour. There is a clear non-linear relationship between hired labour and agricultural output.
Because of the high correlation (0.44) between land surface and hired labour, the effects of the two variables are difficult to disentangle. The farm size is therefore considered as a combination of both.[8] In Figure 6(a), we define the scale of the farm by the respective quantiles of hired labour and land surface used for production. A scale of 0 (resp. 1) means that the farm uses the minimum (resp. maximum) level of hired labour (larger than 0) and the minimum (resp. maximum) surface for production found in the data. Figure 6(a) illustrates that returns to scale of hired labour and land surface are a function of the scale of the farm. Relatively small farms are found to have returns to scale close to 0. Relatively large farms have returns to scale not far below 1. The assumption that returns to scale are not scale dependent - as imposed in the CD model and shown by the horizontal black line - is thus rejected at the 95% confidence interval.

In a second model, we control for land use (see Figure 3). The effects of cultivated land, costs for hired labour and intermediary inputs, and location are similar as for the base model. Farms with a larger share of the farm with banana are found to have a higher agricultural output. The share of coffee planted as the only cash crop positively contributes

[6]The nonparametric model allows for interactions between all regressors. 3-D plots of estimated interactions between regressors are available on request.

[7]We include family labour as an ordered discrete variable with order 10. Results are robust for changing the order to 3 as in the parametric model.

[8]We do not consider the scale effect of intermediary inputs in this analysis because 1) the use of intermediary inputs is not highly correlated to land surface (correlation of 0.12) and 2) the effect of intermediary inputs is insignificant. It should be mentioned that both the physical and economic access to intermediary inputs are rather problematic in the study area.

to production. Again, Figure 6(b) shows that returns to scale are scale dependent.
Model 3 checks for the effects of field characteristics such as the steepness of the plots, perceived soil quality, share of land in marches, application of manure and chemical fertilizers, plot fragmentation (see Figure 4). Steepness of the plots is particularly relevant for this hilly environment. The share of the farm located in the marches is of importance for the production of vegetables. The marches are drained and mostly used for vegetable production. Fragmentation is an important problem. The average number of plots on the farms in the sample is 6.6, with the largest quartile having on average eight plots. We find a non-significant negative effect of steepness of the plots. Fragmentation has a significant non-linear effect at the 90% confidence interval. Perceived soil quality is found to be highly significant. Field characteristics are thus important determinants of agricultural production. The results of the base model concerning the inputs hold. We find a non-linear effect of hired labour on agricultural production and returns to scale that are dependent of the sale of the farm (see Figure 6(c)).
Finally a fourth model checks the effect of off-farm income in total household income, the access to extension services and the age of the head of the farm household (see Figure 5). We do not find significant effects of the three variables. The effect of farm size cultivated is not significant in this model. In contrast to the previous models, we find a significant non-linear positive effect for intermediary inputs in this fourth model. However, as the three added variables are not significant, the model should be interpreted with care. If we drop the three variables, we return to the base model with a significant effect of land surface and a strong non-linear effect of hired labour. Again, model 4 finds that returns to scale are dependent on the scale of the farm (see Figure 6(d)).

In sum, based on this sample of small scale farms, we cannot conclude that it is optimal to concentrate on small farms if the aim is to increase productivity. As RTS are scale dependent, it is possible that unobserved very large farms with low field fragmentation and adapted crop production strategies realize constant or increasing returns to scale. In addition, we find strong effects of crop choice and field characteristics. The agricultural returns from small-scale fragmented production on low soil quality plots are expected to be very low.

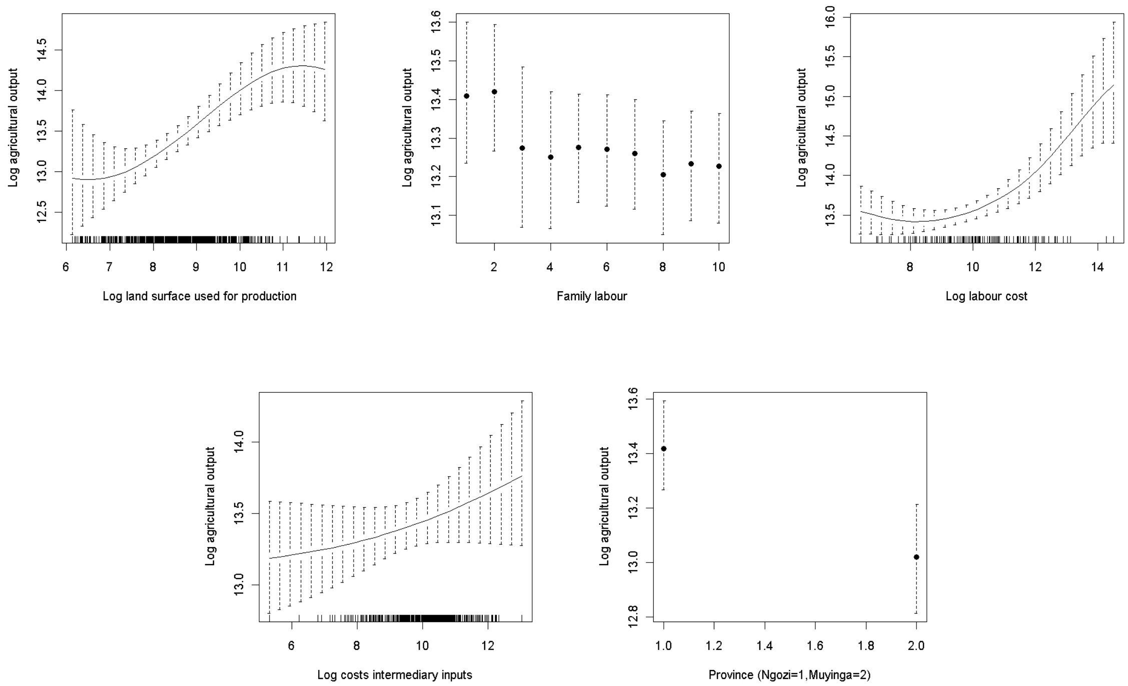


Figure 2: Base Model

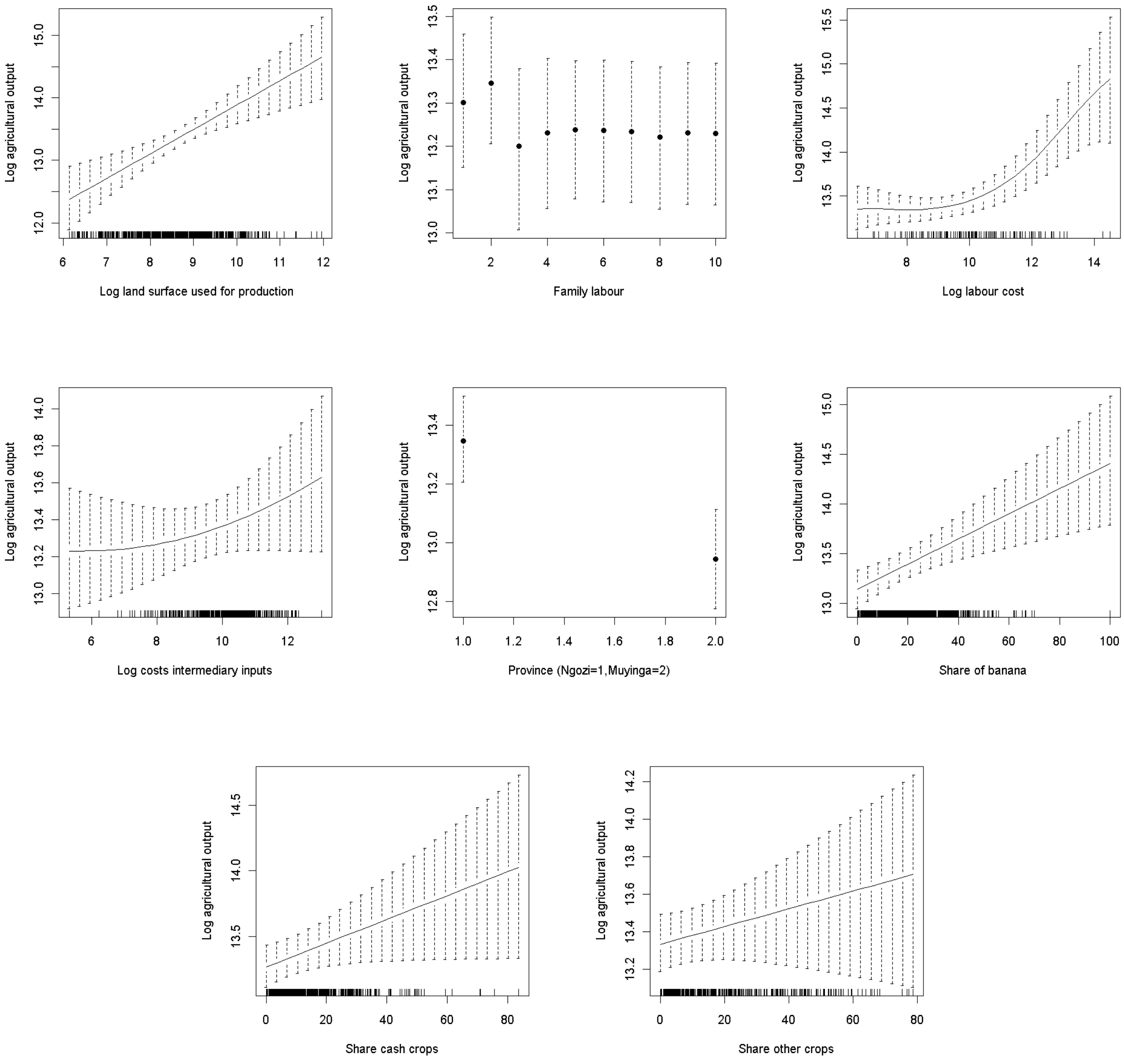


Figure 3: Model 2

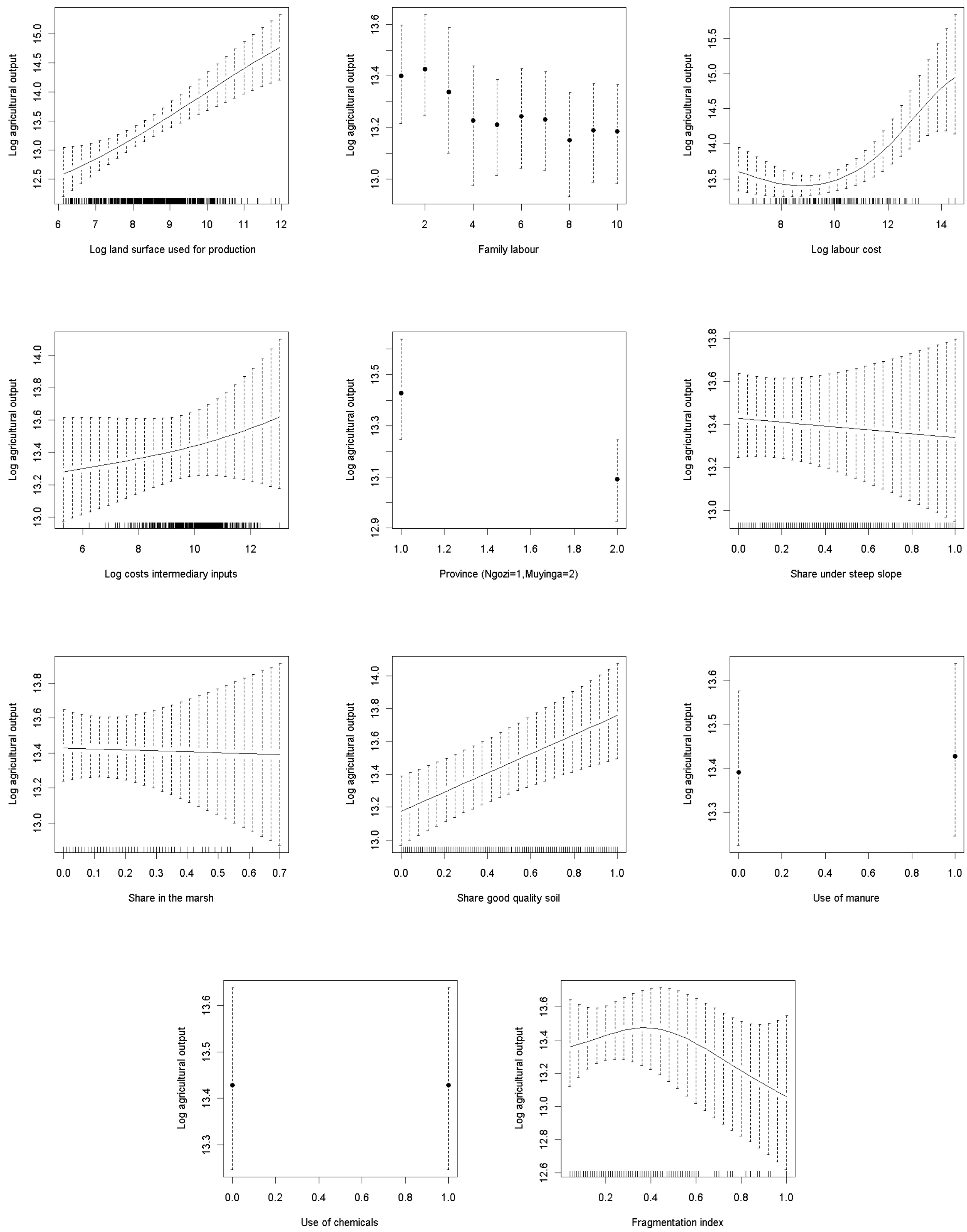


Figure 4: Model 3

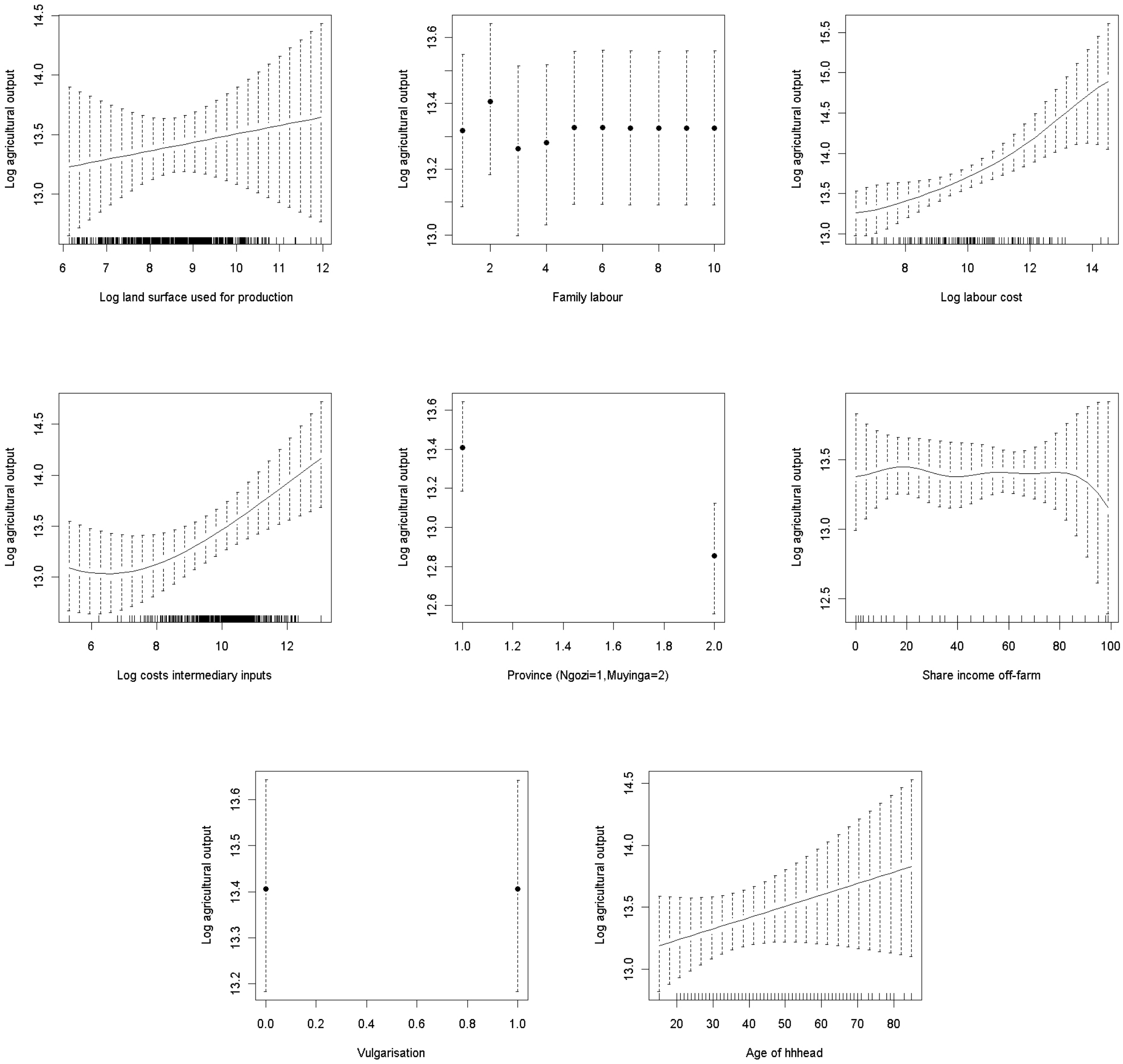
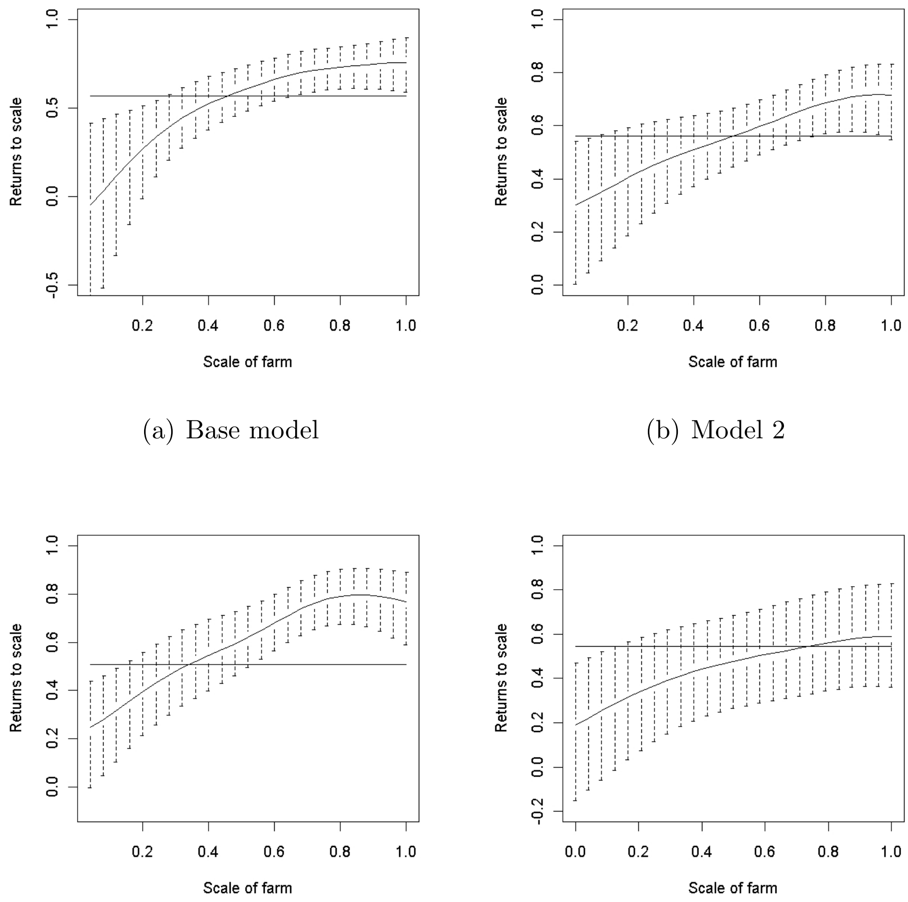


Figure 5: Model 4



(a) Base model

(b) Model 2

(c) Model 3

(d) Model 4

Figure 6: Returns to scale in function of scale of farm

5 Conclusions

Burundi is one of the poorest countries in the world and the farmers interviewed in our research are poor and food insecure. They seem to be trapped in poverty due to a lack of assets, (institutional and economic) access to these assets, and their involvement in low productive activities. Most Burundese households live (partly) from subsistence agriculture. They often lack access to land, and subsequent inheritance and custom risk aversion strategies result in a patchwork of farms that are highly fragmented with mixed cropping patterns. Coffee is their only source of cash from agriculture supplemented with some income from the sales of banana. This said, we find a degree of inequality in land ownership which may explain the importance of land and labour productivity.

Parametric models (Cobb-Douglass and Translog specifications) were not satisfactory to estimate the determinants of crop productivity. We used a nonparametric kernel estimation of the production function (solved with a local-linear estimator) to allow non-linearities and interaction effects. Four different models were estimated controlling for inputs, household, farm and soil characteristics. In each model the effect of size of cultivated land, cost of intermediary inputs and of hired labour was consistent. We find a significant effect of land size and a non-linear effect of hired labour on agricultural output. In addition, crops choice and field characteristics matter. Coffee and banana production are found to yield higher returns compared to the other crops. Fragmentation and low perceived soil quality are associated with low agricultural productivity.

The model confirms that farm size itself matters for the relationship between its size and productivity. Our findings confirm both the relatively high productivity of the very small farms, but it also shows the economies of scale that larger farms may exploit. This is a confirmation of the comments made in Dercon and Collier (2009) on the farming scales that are compared in IR literature, namely that the range of farm sizes studied with parametric econometric models is not large enough to show the true relationship between size and productivity. Our results confirm that the effect of size on production is different over the size spectrum. Hence, the potential contribution of agriculture to the potential improvement of the households' livelihoods is different. The implication for policy makers should be to rethink their focus on smallholder agriculture. The options for diversification out of agriculture for these small farms are rather small and they are limited to low paid irregular jobs on other peoples farms or businesses. Yet exploring new better-paid and protected rural non-farm opportunities for the smallest farms is an area for further research. Another topic that we want to explore in the near future are the possible agricultural policy options for optimizing farm production. This includes possibilities for exploiting economies of scale by crop specialization and reducing land fragmentation.

References

Aitchision, J., Aitken, C. G. G., 1976. Multivariate binary discrimination by kernel method. Biometrika 63 (3), 413–420.

Assunção, J., Braido, L., 2007. Testing household-specific explanations for the inverse productivity relationship. American Journal of Agricultural Economics 89 (4), 980–990.

Assunção, J., Ghatak, M., 2003. Can unobserved heterogeneity in farmer ability explain the inverse relationship between farm size and productivity. Economics Letters 80, 189–194.

Barrett, C., 1996. On price risk and the inverse farm size-productivity relationship. Journal of Development Economics 51, 193–215.

Barrett, C., Bellemare, M., Hou, J., 2010. Reconsidering conventional explanations of the inverse productivity-size relationship. World Development 38, 88–97.

Benjamin, D., 1995. Can unobserved land quality explain the inverse productivity relationship? Journal of Development Economics 46, 51–85.

Boserup, E., 1965. The Conditions of Agricultural Growth: The Economics of Agrarian Change under Population Pressure. Allen & Unwin: London.

Bundervoet, T., 2009. Livestock, land and political power: The 1993 killings in burundi. Journal of Peace Research 46, 357–376.

Byiringiro, F., Reardon, T., 1996. Farm productivity in rwanda: effects of farm size, erosion, and soil conservation investments. Agricultural Economics 15, 127–136.

CIA, 2010. The world factbook.
URL `https://www.cia.gov/library/publications/the-world-factbook/`

Cochet, H., 2004. Agrarian dynamics, population growth and resource management : The case of burundi. GeoJournal 60, 111–122.

Collier, P., Dercon, S., 2009. African agriculture in 50 years: Smallholders in a rapidly changing world? In: FAO, UN Economic and Social Development Department.

Dercon, S., Krishnan, P., 1996. Income portfolios in rural ethiopia and tanzania: Choices and constraints. Journal of Development Studies 32, 850–875.

FAO, 2009. Food insecurity in the world, economic crises impacts and lessons learned. FAO (Food and Agriculture Organisation): Rome.

Feder, G., 1985. The relation between farm size and farm productivity : The role of family labor, supervision and credit constraints. Journal of Development Economics 18, 297–313.

Hall, P., Li, Q., Racine, J. S., Nov. 2007. Nonparametric estimation of regression functions in the presence of irrelevant regressors. Review Of Economics And Statistics 89 (4), 784–789.

Heltberg, R., 1998. Rural market imperfections and the farm size-productivity relationship: Evidence from pakistan. World Development 26, 1807–1826.

Henderson, D. J., Kumbhakar, S. C., Jul. 2006. Public and private capital productivity puzzle: A nonparametric approach. Southern Economic Journal 73 (1), 219–232.

Jayne, T. S., Yamano, T., Weber, M. T., Tschirley, D., Benfica, R., Chapoto, A., 2003. Smallholder income and land distribution in africa: implications for poverty reduction strategies. Food Policy 28, 253–275.

Lamb, R., 2003. Inverse productivity: Land quality, labor markets, and measurement error. Journal of Development Economics 71, 71–95.

Leathers, H., Foster, P., 2009. The World Food Problem. Toward Ending Undernutrition in the Third World. Fourth Edition. Lynne Rienner Publishers: Boulder, Colorado.

Li, Q., Racine, J., Apr. 2004. Cross-validated local linear nonparametric regression. Statistica Sinica 14 (2), 485–512.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: theory and practice. Princeton University Press.

Lipton, M., 2010. From policy aims and small-farm characteristics to farm science needs. World development 10, 1399–1412.

Piesse, J., Thirtle, C., 2010. Agricultural r&d, technology and productivity. Philosophical transactions of the Royal Society B 365, 3035–3047.

Platteau, J., 1996. The evolutionary theory of land rights as applied to sub-saharan africa: A critical assessment. Development and Change 27, 29–86.

Racine, J., Li, Q., Mar. 2004. Nonparametric estimation of regression functions with both categorical and continuous data. Journal Of Econometrics 119 (1), 99–130.

Rishirumuhirwa, T., Roose, E., 1998. The contribution of banana farming systems to sustainable land use in burundi. Advances in GeoEcology 31, 1197–1204.

Schultz, T., 1964. Transforming Traditional Agriculture. Yale University Press: New Haven.

Wang, M. C., van Ryzin, J., 1981. A class of smooth estimators for discrete-distributions. Biometrika 68 (1), 301–309.

Wiggins, S., Kisten, J., Llambi, L., 2010. The future of small farms. World 38, 1341–1348.

WorldBank, 2009. World development indicators.
URL `http:`
`www.worldbank.org`